DOCUMENT RESUME

ED 100 591 RC 008 283

AUTHOR	Giesecke, G. Lee
TITLE	The Use of Spanish Surnames as a Means of Identifying Latinos in the United States and Puerto Rico. Professional Paper No. 22-74.
INSTITUTION	Human Resources Research Organization, Alexandria, Va.
SPONS AGENCY	Office of the Assistant Secretary of Defense for Manpower and Reserve Affairs (DOD), Washington, D.C.
REPORT NO	HumRRO-PP-22-74
PUB DATE	Dec 74
NOTE	23p.; For related document, see ED 085 146
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Classification; *Codification; Comparative Analysis; *Computer Science; *Information Processing; Input Output Analysis; *Spanish Speaking; Tables (Data)
IDENTIFIERS	*Names

ABSTRACT

Responses from 1,613 former servicemen with Spanish surnames were examined to evaluate Spanish surname recognition as a means of identifying persons of Latin American birth or ancestry. Questionnaires were mailed to 5,019 servicemen who had: (1) been inductees or enlistees in the Army, Navy, or Marine Corps; (2) left active duty in the first 6 months of 1972; (3) a Reserve military obligation; and (4) a Spanish surname according to one or more procedures. Procedures determining whether a name was Spanish were all computerized. All but one depended on sorting the names alphabetically and comparing each with lists of Spanish surnames. The remaining procedure, developed by Buechley, depended on letter combinations and surname endings. The lists used were: (1) census surnames, (2) "broad" Spanish surnames, (3) "narrow" Spanish surnames, and (4) Morton's Spanish surnames. Although a determination was made via computer as to whether each person had a "narrow" Spanish first name, this fact was used only in the analysis. Some findings were: (1) outside certain areas Spanish surname recognition included a high proportion of persons who did not meet other criteria indicating Latin American birth or ancestry; and (2) within five Southwestern states, increasing proportions of persons who could not be classified as Spanish at increasing education and aptitude levels were included. (NQ)

SCOPE OF INTEREST NOTICE
RC UD
30

Professional Paper 22-74

HumRRO-PP-22-74

HumRRO

BEST COPY AVAILABLE

RECEIVED FEB 3 1975

The Use of Spanish Surnames As a Means of Identifying Latinos in The United States and Puerto Rico

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

G. Lee Giesecke

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

ED100591

Approved for public release; distribution unlimited.

December 1974

Prepared for

Directorate for Manpower Research
Office of the Assistant Secretary of Defense
(Manpower and Reserve Affairs)
Department of Defense

RC008283

0002

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development, and consultation.

The contents of this publication do not necessarily represent the official opinion or policy of the sponsor of the HumRRO research.

Published
December 1974
by
HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street
Alexandria, Virginia 22314

0003

Prefatory Note

This research was conducted under the sponsorship of the Assistant Secretary of Defense (Manpower and Reserve Affairs). The analysis was performed by G. Lee Giesecke, Senior Technician, HumRRO Division No. 7, Human Resources Research Organization, Alexandria, Virginia.

THE USE OF SPANISH SURNAMES AS A MEANS OF IDENTIFYING LATINOS IN THE UNITED STATES AND PUERTO RICO

G. Lee Giesecke

INTRODUCTION

Despite the heterogeneity of persons living in the United States with Latin-American birth or ancestry, Latinos (as I will call this group) share a number of distinguishing characteristics. Among these features—which are not all present in all cases—are Catholic faith, Spanish language, a Spanish surname, derivation from countries of similar history and culture, and membership in an unassimilated and, in some cases, disadvantaged minority. In part because of these similarities and in part as a political and statistical expedient, there is some tendency to treat Latinos as a single composite, a tendency that has led to a number of contradictions.

Latinos are sometimes referred to as "Spanish-surnamed" individuals, as if the names rather than the cultures were the important group-defining characteristic (1). On the other hand, if Spanish language is used to define the group, we are led to self-contradictory statements like that of Ramirez who informs us that "all . . . (Spanish-speaking persons) . . . do not speak Spanish" (2). It is, of course, not necessary to live with such contradictions. Any study sufficiently restricted in size and geography can afford to be culture specific. Where these conditions are not met, however, it is generally necessary to define a Spanish group by a somewhat arbitrary characteristic such as Spanish language or the presence of a Spanish surname. In this paper, the identification of Latinos by the use of Spanish surnames is examined, along with the problems inherent in the use of this method and some possible alternatives to it.

DATA

The data are taken from responses to the questionnaire shown in Appendix A. The questionnaires were mailed to a sample of 5,019 former servicemen, all of whom had been inductees or enlistees in the Army, Navy, or Marine Corps, had left active duty in the first six months of 1972, had a Reserve military obligation, and had a Spanish surname according to one or more of several procedures.

The procedures used to determine whether a name was Spanish were all computerized. All but one depended upon sorting the names to be classified alphabetically and comparing each with lists of Spanish surnames, which were also sorted alphabetically. The remaining procedure for classifying names as Spanish was developed by Buechley and depends on letter combinations and surname endings (3). This latter technique will be referred to as the Buechley technique. The other procedures will be identified by the surname list that was used to make the classification. The surname lists are as follows:

(1) Census surnames—the list used in the 1970 Census to identify persons with Spanish surnames in the southwestern United States (4).[1]

[1] In addition to the names on the list cited here, the 1970 Census classified as Spanish names prefixed by "De," "Del," "De La," "De Las," or "De Los." Also the name "Martin," although on the list, was classified by the Census as Spanish only if certain other characteristics were present. In both respects the 1970 Census differed from the present study. It also differed in a third respect, in that the name classification was performed manually, rather than by computer.

(2) Broad Spanish surnames—a list that contained additions to the 1970 Census list.[1]

(3) Narrow Spanish surnames—a subset of the broad surname list.[1] This list excluded certain names that were likely to be culturally ambiguous in the United States.

(4) Morton's Spanish Surnames—a list prepared by Dr. William E. Morton (6).

In addition, a determination was made via computer as to whether each person had a "narrow" Spanish first name, although this fact was used only in the analysis. No one received a questionnaire solely because he had a Spanish first name.

Items 1 and 4 on the questionnaire were adopted from the 1970 census. Item 3 appears as it did on the Airman Sample Survey of March and July of 1971 with the exception that a space was added in the present study for specifying the "other" category (7). The remaining items explore areas not covered by these three questions.

SURVEY RESPONSE

The stratification of the sample and responses to the survey are given in Table 1. Although overall response rate to the survey was only 32%, there was no indication of response bias by geographic areas ($\chi^2 = 4.3$, $3df$), by branch of service ($\chi^2 = 2.2$, $2df$) or by whether the individual was an inductee instead of an enlistee ($\chi^2 = 1.2$, $1df$). There was also no indication of a response bias depending upon whether or not an individual had a "narrow" Spanish surname ($\chi^2 = 0.3$ $1df$) or a "narrow" Spanish first name ($\chi^2 = 0.5$, $1df$). There was a clear response bias by educational level ($\chi^2 = 75.4$, $2df$). However, this had been anticipated in the original stratification of the sample and is handled by appropriate weighting.

When the respondents were separated into three groups according to how quickly their questionnaires were returned, there was no significant difference between groups in the proportion of respondents where Spanish was spoken in the home when they were children ($\chi^2 = 4.1$, $2df$).

Among those who did respond to the survey, there were relatively few unknown items. The numbers and percentages of unknowns are given on an item-by-item basis in Table 2.

TABULATIONS

Except where otherwise specified, all the tabulations conform to the following conventions:

(1) Cases with an unknown item are excluded for any cell in a tabulation requiring the item, but not necessarily for the entire table.

(2) Numbers are presented as unweighted.

(3) Percentages are computed by weighting the individual cases to reflect the population shown in Table 1.

For Items 1, 2, 3, 9, and 10 the results reflect anything written in the "specify" space on the questionnaire as well as the marked response. In tables where there is a Spanish/non-Spanish dichotomy, responses "a" through "d" on Question 2, "a" through "e" on Question 4, and "a" and "g" on Question 9 are treated as Spanish. In cases where

[1] The lists of "broad" Spanish surnames, "narrow" Spanish surnames, and "narrow" Spanish first names are described by Giesecke (5).

0006

Table 1

Population Sample, Responses, and Response Rates, By Geographic Area and Educational Level

BEST COPY AVAILABLE

Years of School Completed	Geographic Area				
	Five Southwestern States[a]	New York, New Jersey, Florida	Puerto Rico	Other Areas	All Areas
Less than 12					
Population	1,657	584	189	919	3,349
Sample	521	277	89	225	1,112
Responses	132	74	28	46	280
Percent Responding	25	27	31	20	25
12					
Population	4,550	1,153	644	2,995	9,342
Sample	562	517	350	1,025	2,454
Responses	182	154	117	291	744
Percent Responding	32	30	33	28	30
Over 12					
Population	1,295	476	183	1,376	3,330
Sample	649	162	94	548	1,453
Responses	262	63	39	225	589
Percent Responding	40	39	42	41	41
All Educational Levels					
Population	7,502	2,213	1,016	5,290	16,021
Sample	1,732	956	533	1,798	5,019
Responses	576	291	184	562	1,613
Percent Responding	33	30	35	37	32

[a]Arizona, California, Colorado, New Mexico, Texas

more than one answer was coded, the response was treated as Spanish if one of the responses was Spanish. However, for Question 10, only the first response was used. If it were anthing but "h", "i", "m", or "p" the person was classified as Spanish.

RESULTS

Cases Misclassified as Spanish

For each of the surname classification procedures, the extent to which persons with Spanish surnames do not meet various other criteria is shown in Table 3. As was shown by earlier studies (5,8) and presented in Table 3, a substantial proportion of persons with Spanish surnames cannot be classified as Spanish by certain other criteria. It is possible to bypass this problem to some extent by using a surname list with fewer culturally ambiguous names, as with the "narrow" surname list used here. However, the results are still not completely satisfactory when the United States is taken as a whole.

0007

Table 2

Number and Percent of "Unknowns" for the Survey Items

Questionnaire Item	Unknowns (N)	Percent Unknown[a]
1. Language spoken in childhood home	5	0.3
2. Current use of Spanish	11	0.7
3. Self identity	4	0.2
4. Origin or descent	12	0.7
5. Parents born in Spanish-speaking country or territory	7	0.4
6. Grandparents born in Spanish-speaking country or territory	14	0.9
7. Parents fluent in Spanish	17	1.1
8. Grandparents fluent in Spanish	30	1.9
9. Origin or nationality of last name	86	5.3
10. Have most in common with. . .	17	1.1

[a]The percentage is not weighted.

Curiously, Table 3 presents a more optimistic picture for Spanish surname recognition than do earlier reports. Giesecke showed that for 917 Air Force enlisted men who had Spanish surnames according to the 1970 Census list (matched via computer), 38% did not classify themselves as "Spanish or Mexican-American" (5). A similar figure of 37% is found using the present data by not counting as Spanish persons who have marked "other" and specified some other Spanish group on Item 3 of the questionnaire. When the latter cases are properly identified as "Spanish," the percent misclassified as Spanish using Census surnames drops to 30%.

Several other criteria in Table 3 present a still more favorable picture of Spanish surname recognition. The origin or descent criterion (Item 4) is not met by 21% of persons whose surnames were recognized (via computer) as being on the 1970 Census list. This figure is lower than an estimate based on the March 1971 Current Population Survey (CPS) in which 38% of husbands in husband-wife households with Spanish surnames (using the Census list, with manual coding) were not of Spanish origin or descent (8). Although the figures in the CPS report cannot be reconciled with those of the present report, it is clear from both studies that Spanish surname recognition leaves something to be desired when the United States is taken as a whole.[1]

[1] The studies differ in the surname recognition techniques, the populations involved, and the phrasing of the "origin or descent" question. Apparently the phrasing of the "origin or descent" question does not make too much of a difference or the CPS projected population of Spanish origin would not be so close to that of the 1970 Census, which differed in the phrasing of this question. Unfortunately, data are not available to adjust for the population differences between the present study and the CPS study. However, the present study population has higher educational levels and should as a consequence have a larger proportion misclassified as Spanish, despite the presence of the small Puerto Rican group. This leaves the differences in the name recognition techniques as the most likely source of the differences.

0008

Table 3

Number and Percent With Surnames Classified as Spanish Who Were Not Spanish by Various Other Criteria

BEST COPY AVAILABLE

Questionnaire Item	Criterion	Surname Classification Procedure[a]									
		"Narrow"		"Broad"		Census		Morton		Buechley	
		N	%	N	%	N	%	N	%	N	%
1	Spanish Spoken in Childhood Home	119	11.5	327	24.7	270	23.0	292	23.1	469	30.6
2	Some Spanish Spoken Currently	157	16.9	361	28.7	305	27.4	329	27.5	507	34.8
3	Considers Self as Spanish or Mexican American	210	19.6	428	31.9	360	30.3	389	30.5	580	37.9
4	Spanish Origin or Descent	90	9.3	295	22.8	242	21.3	259	20.9	444	29.7
5	Either Parent Born in Spanish-Speaking Country or Territory	318	37.7	536	47.0	472	47.0	498	46.5	692	52.1
6	Two or More Grandparents Born in Spanish-Speaking Country or Territory	191	22.0	403	33.5	346	33.1	369	32.7	549	39.1
7	Either Parent Could Speak Spanish	86	8.1	286	21.6	235	20.1	253	19.9	427	27.7
8	Two or More Grandparents Could Speak Spanish	171	18.8	380	31.1	315	29.3	334	29.2	516	36.1
9	Considers Name to be Spanish or Mexican	92	8.8	277	21.6	224	19.8	245	19.9	429	28.9
10	Had Most in Common With a Spanish Group	217	22.3	415	32.4	352	31.0	379	31.3	558	38.4
—	Had Narrow Spanish First Name	572	62.1	800	67.4	706	66.8	751	67.9	955	71.0
—	Spanish by Three or More Survey Criteria	73	7.0	277	20.6	227	19.1	244	18.8	421	27.1
—	Spanish According to Discriminant Function[b]	84	8.0	291	21.8	237	20.1	257	20.1	434	28.0

[a]The numbers are the unweighted number who did not meet the criterion but whose names were classified as Spanish by a particular surname classification procedure. The denominator for each percentage is the number treated as Spanish by a particular surname classification procedure. Prior to caculating the percentages, the cases are weighted to reflect the population shown in Table 1.

[b]The discriminant function was set to distinguish cases that were Spanish by three or more survey criteria.

Culturally Ambiguous Names. Of special interest in Table 3 is the large number who did not consider their names to be Spanish. Of those who had "narrow" Spanish surnames, 91 (or 9%) did not classify their names as Spanish or Mexican on the questionnaire. Not only is acculturation a consideration in the sense that some persons with Spanish names can no longer be considered Latinos, but culturally ambiguous names are still a problem.

The problem of culturally ambiguous names is demonstrated in Table 4 in greater detail for persons with "narrow" Spanish surnames. About 62% of the persons who classified their names as something besides Spanish or Mexican would be classified as Spanish by less than three of the other criteria. These persons are best classified as non-Spanish, since it is relatively easy to appear Spanish by one or two of the survey criteria by happenstance circumstances.

The nationalities of names most commonly misclassified as Spanish are Portuguese, Italian, French, and Filipino, although a variety of other nationalities are involved as well. The chief cause of these overlapping names is the historical or geographical links between Spain and these other nations. However, some of the ambiguous names appear to have arisen spontaneously from more than one source.

Table 4

Number With Narrow Spanish Surnames by Self-Classification of Surname and Further Survey Criteria by Which They Might be Considered Spanish

Self-classification of Surname	Number of Further Survey Criteria by Which Persons Might be Classified as Spanish										
	0	1	2	3	4	5	6	7	8	9	All Cases
Spanish	4	8	2	4	11	27	32	96	145	256	585
Mexican	2	0	0	1	2	6	11	44	69	81	216
Portugese	13	4	1	4	0	1	0	0	2	1	26
French	4	3	0	0	0	0	2	2	4	7	22
Italian	13	0	0	0	1	0	0	1	1	3	19
Filipino	4	1	1	1	0	0	0	0	0	0	7
German	5	0	0	0	0	0	0	0	0	0	5
Other	8	0	0	0	0	0	1	1	1	2	13
Not Known	2	0	0	0	1	2	3	7	8	12	35

Results by Geographic Area. When broken out geographically, the percentage of cases misclassified as Spanish approaches acceptable levels (by some criteria) if restricted to Puerto Rico; New York, New Jersey, Florida; and five southwestern states, as shown in Table 5. No doubt, there are also local areas outside these states where Spanish surname recognition is also acceptable, as well as areas within these states where the correspondence between Spanish name and Spanish culture is poor. Nevertheless, the probability of making a classification error clearly depends upon geographic location.

Results by Other Characteristics. More generally, a good case can be made that the validity of Spanish surname recognition varies with any variable associated with acculturation. This is demonstrated for educational level in Table 6 and percentile score on the Armed Forces Qualification Test (AFQT) in Table 7. Both tables are restricted to ex-servicemen living in the five southwestern states.

Interestingly, as shown in Tables 6 and 7, the Census surnames degrade more in specificity at higher educational and AFQT levels than do the "narrow" surnames. This larger misclassification gradient is undoubtedly caused by the larger proportion of culturally ambiguous names included on the Census surname list.

0010

Table 5

Number and Percent With Surnames Classified as Spanish Who Were Not Spanish by Other Criteria, by Geographic Area and Surname Recognition Technique

BEST COPY AVAILABLE

Criterion	Questionnaire Item	Geographic Area and Surname Recognition Technique[a]															
		Five Southwestern States[b]				New York, New Jersey, Florida				Puerto Rico				Elsewhere			
		"Narrow"		Census		"Narrow"		Census		"Narrow"		Census		"Narrow"		Census	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Spanish Spoken in Childhood Home	1	41	7.6	55	9.8	17	11.5	42	25.9	3	1.8	3	2.0	58	33.5	170	59.9
Some Spanish Spoken Currently	2	68	13.5	83	15.9	17	11.6	42	26.1	0	0.0	0	0.0	72	44.1	180	64.4
Considers Self as Spanish or Mexican American	3	50	10.3	64	12.7	31	21.2	54	33.3	51	29.8	49	32.5	68	41.2	180	65.0
Spanish Origin or Descent	4	41	7.8	58	10.9	14	9.6	36	22.3	0	0.0	0	0.0	47	26.7	159	55.6
Either Parent Born in Spanish-Speaking Country or Territory	5	220	45.5	233	47.8	12	8.2	39	24.1	8	4.5	8	5.1	78	48.0	192	69.3
Two or More Grandparents Born in Spanish-Speaking Country or Territory	6	115	24.4	132	27.2	12	8.3	37	23.2	9	5.1	7	4.5	55	32.6	170	60.7
Either Parent Could Speak Spanish	7	22	4.2	36	6.7	10	6.8	36	22.2	7	4.0	5	3.3	47	26.7	158	56.0
Two or More Grandparents Could Speak Spanish	8	63	15.3	77	17.5	30	21.0	52	32.5	23	13.8	18	12.1	55	34.1	168	61.6
Considers Name to be Spanish or Mexican	9	27	4.8	46	8.5	15	10.6	36	23.4	7	4.2	5	3.1	43	25.2	137	50.9
Had Most in Common With a Spanish Group	10	103	19.1	118	21.5	36	24.6	58	35.7	6	3.4	6	3.9	81	49.1	184	64.7
Had Narrow Spanish First Name	--	301	63.1	310	63.3	77	52.6	99	61.0	79	47.9	71	49.0	115	72.7	226	83.2
Spanish by Three or More Survey Criteria	--	20	3.5	36	6.2	8	5.4	33	20.2	0	0.0	0	0.0	45	25.9	158	55.5
Spanish According to Discriminant Function[c]	--	25	4.6	40	7.2	9	6.1	35	21.4	0	0.0	0	0.0	50	28.1	162	57.0

[a]The numbers are the unweighted number who did not meet the criterion but whose names were classified as Spanish by a particular surname classification procedure. The denominator for each percentage is the number treated as Spanish by a particular surname classification procedure. Prior to calculating the percentages, the cases are weighted to reflect the population shown in Table 1.

[b]Arizona, California, Colorado, New Mexico, Texas.

[c]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

0011

Table 6

Number and Percent From Five Southwestern States With Spanish Surnames but Not Spanish by Other Criteria, by Educational Level and Surname Recognition Technique

Criterion	Questionnaire Item	Years of School Completed and Surname Recognition Technique[a]											
		Less Than 12				12				Over 12			
		"Narrow"		Census		"Narrow"		Census		"Narrow"		Census	
		N	%	N	%	N	%	N	%	N	%	N	%
Spanish Spoken in Childhood Home	1	1	0.8	2	1.7	13	8.5	16	10.7	27	14.5	37	19.0
Some Spanish Spoken Currently	2	6	5.1	7	5.9	23	15.0	26	17.3	39	21.0	50	25.6
Spanish Origin or Descent	4	5	4.3	8	6.9	12	7.9	16	10.8	24	13.0	34	17.6
Considers Name to be Spanish or Mexican	9	4	3.4	5	4.2	6	4.2	12	8.5	17	9.4	29	15.4
Spanish According to Discriminant Function[b]		0	0.0	1	0.8	8	5.2	12	8.0	17	9.1	27	13.8

[a]The numbers are the unweighted number who did not meet the criterion but whose names were classified as Spanish by a particular surname classification procedure. The denominator for each percentage is the number treated as Spanish by a particular surname classification procedure. Prior to calculating the percentages, the numbers are weighted to reflect the population shown in Table 1. The five Southwestern states were Arizona, California, Colorado, New Mexico, Texas.

[b]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

Also of interest in Tables 6 and 7 are the smaller misclassification gradients displayed by the "origin or descent" (Item 4) and "considers name Spanish" (Item 9) criteria. Apparently the gradients are caused in part by acculturation and in part by culturally ambiguous names. The fact of speaking Spanish or having a Spanish mother tongue is more likely to be lost through acculturation than the facts of being of Spanish descent or having a Spanish surname.

Cases Misclassified as Non-Spanish

Because all the persons sampled had a Spanish surname by at least one classification procedure, we cannot obtain in this study a true percentage of persons who met a particular criterion but were missed by a particular surname recognition technique. Nevertheless, it is possible to compare the relative completeness rates of different surname techniques, so long as it is understood that the present study can address itself only to a subset of the population who might meet a particular criterion.

In Table 8 are shown the cases that could be classified as Spanish by various survey criteria but were missed by particular surname classification procedures. Except for survey Item 5, the Buechley technique is shown to be the most complete. This same result was found in the earlier Giesecke study (5).

Table 7

Number and Percent From Five Southwestern States With Spanish Surnames But Not Spanish by Other Criteria, by Surname Recognition Technique and Percentile Score on the Armed Forces Qualification Test (AFQT)

BEST COPY AVAILABLE

Criterion	Questionnaire Item	AFQT Percentile and Surname Recognition Technique[a]											
		1-33				34-67				68-100			
		"Narrow"		Census		"Narrow"		Census		"Narrow"		Census	
		N	%	N	%	N	%	N	%	N	%	N	%
Spanish Spoken in Childhood Home	1	10	5.2	10	5.7	12	6.7	10	5.5	18	14.3	33	25.7
Some Spanish Spoken Currently	2	16	10.7	17	11.6	22	13.0	20	11.9	29	21.4	44	32.5
Spanish Origin or Descent	4	10	5.9	12	7.4	14	8.6	14	8.7	16	9.5	30	20.9
Considers Name to be Spanish or Mexican	9	8	3.6	8	4.5	8	6.0	12	8.0	11	5.7	25	18.2
Spanish According to Discriminant Function[b]	–	6	2.6	6	3.0	9	6.1	8	5.9	9	4.3	24	16.2

[a]The numbers are the unweighted number who did not meet the criterion but whose names were classified as Spanish by a particular surname classification procedure. The denominator for each percentage is the number treated as Spanish by a particular surname classification procedure. Prior to calculating the percentages, the numbers are weighted to reflect the population shown in Table 1. The five Southwestern states are Arizona, California, Colorado, New Mexico, Texas.

[b]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

Using the discriminant function as the criterion, the superior completeness of the Buechley technique is demonstrated for most population subsets, as shown by Table 9. The Buechley technique, however, is not 100% complete and cannot be, since a number of persons of Spanish culture do not have Spanish surnames. Giesecke estimates for an Air Force population that the Buechley technique would miss at least 8% of persons who might classify themselves as "Spanish or Mexican-American" (5).

Since 100% completeness appears to be out of the question, it is the author's view that a restricted or "narrow" list of Spanish surnames should be used for most statistical purposes, if Spanish surname recognition is used at all. As shown in Tables 8 and 9, the penalty in completeness is not severe, and as shown in the earlier table, the reduced proportion of cases misclassified as Spanish can be important.

Alternatives to Spanish Surname Recognition

So far we have shown that outside certain areas the "narrow" surname technique leads to an unacceptably high proportion of cases misclassified as Spanish. We have shown also that, even within the five southwestern states, the proportion of cases misclassified as Spanish increases substantially at higher educational and aptitude levels. It has been shown elsewhere that Spanish surname recognition is far from complete (5,8). Presumably one can find alternatives to Spanish surname recognition that do not have

Table 8

Number and Percent of Persons Who Could Be Classified as Spanish According to Particular Criteria but Whose Surnames Were Not Classified as Spanish

BEST COPY AVAILABLE

Criterion	Questionnaire Item	Surname Classification Procedure[a]									
		"Narrow"		"Broad"		Census		Morton		Buechley	
		N	%	N	%	N	%	N	%	N	%
Spanish Spoken in Childhood Home	1	77	9.2	40	5.1	114	12.8	86	8.4	35	3.4
Some Spanish Spoken Currently	2	79	9.8	38	5.0	114	13.2	89	8.9	42	4.2
Considers Self as Spanish or Mexican-American	3	54	7.7	25	4.0	89	11.7	68	7.6	29	3.3
Spanish Origin or Descent	4	78	8.9	39	4.6	118	12.8	85	7.9	41	4.0
Either Parent Born in Spanish-Speaking Country or Territory	5	53	9.0	26	4.9	93	14.5	70	9.4	36	5.1
Two or More Grandparents Born in Spanish Speaking Country or Territory	6	69	9.7	36	5.3	110	14.4	84	9.5	37	4.3
Either Parent Could Speak Spanish	7	91	9.9	48	5.8	129	13.9	98	9.3	46	4.1
Two or More Grandparents Could Speak Spanish	8	73	9.4	39	5.5	102	12.9	76	8.5	29	3.2
Considered Name to be Spanish or Mexican	9	69	8.6	36	4.9	102	11.8	76	7.6	29	3.2
Had Most in Common With a Spanish Group	10	90	11.5	47	5.9	116	13.9	94	9.9	41	5.1
Had Narrow Spanish First Name	--	30	8.3	12	3.0	48	11.3	43	10.0	19	4.7
Spanish by Three or More Survey Criteria	--	88	9.7	46	5.3	128	13.4	95	8.6	42	3.9
Spanish According to Discriminant Function[b]	--	82	9.3	43	5.2	121	13.1	91	8.6	38	3.6

[a]The numbers are the unweighted number who met the criterion but were not classified as Spanish by a particular surname classification procedure. The denominator for each percentage is the number meeting the criterion. Prior to calculating the percentages, cases were weighted to reflect the population shown in Table 1.

[b]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

Table 9

Number and Percent Classified as Spanish by the Discriminant Function With Surnames Not Classified as Spanish, by Population Subset and Surname Recognition Technique[a]

BEST COPY AVAILABLE

Population Subset	Surname Recognition Technique									
	"Narrow"		"Broad"		Census		Morton		Buechley	
	N	%	N	%	N	%	N	%	N	%
Geographic Area										
Five Southwestern States	34	7.8	18	4.5	42	10.6	25	5.6	7	1.7
New York, New Jersey, Florida	19	12.2	12	7.6	29	18.5	27	17.3	12	7.6
Puerto Rico	10	5.3	2	1.0	31	16.8	26	14.5	11	6.0
Elsewhere	19	16.4	11	9.5	19	17.4	12	11.0	8	6.9
Years of School Completed										
Under 12	10	7.8	10	4.6	10	7.8	28	10.5	9	3.8
12	13	8.2	23	5.6	20	12.7	46	8.1	21	3.5
Over 12	11	6.1	10	4.2	12	6.6	17	7.1	8	3.4
AFQT Percentile										
1-33	12	7.4	14	3.2	15	10.6	50	9.3	25	5.5
34-67	15	9.1	17	7.1	14	9.7	23	7.3	4	0.9
68-100	6	6.5	8	6.1	12	13.7	13	8.9	6	3.0

[a]The numbers are unweighted. The percentages are weighted to reflect the population shown in Table 1. Prior to weighting all of the denominators included at least 100 cases. The discriminant function was set to distinguish cases that were Spanish by three or more criteria.

these disadvantages. Unfortunately, using the present data we cannot test the adequacy of alternatives to Spanish surname recognition, since our population is restricted to persons of Spanish surnames. For this limited population we can, however, show the percentage of persons who met a particular survey criterion but failed to meet a second criterion. This is done in Table 10.

Judging from the columns of Table 10, the most complete survey items are Item 1 (Spanish mother tongue), Item 4 (Spanish origin or descent), Item 7 (either parent spoke Spanish), and Item 9 (considers name Spanish). As might be expected, Item 2 (some Spanish spoken currently), Item 5 (either parent born in Spanish-speaking country or territory), and Item 6 (two or more grandparents born in Spanish-speaking country or territory) were somewhat less inclusive. The Spanish identity questions (Items 3 and 10) proved to be less inclusive than even "Spanish Spoken Currently" (Item 2). This same ordering of the survey items in terms of inclusiveness is shown in Table 3.

Three of the survey items that were relatively inclusive for the survey population, Items 1, 4, and 7, are shown in Table 10 to be relatively specific as well. This may be seen by comparing the rows of Table 10. Only three survey items (2, 3, and 5) appear to do as well in terms of specificity.

The specificity of the two Census items (Survey Items 1 and 4) can be seen on a more comprehensive population from an earlier Current Population Survey (CPS) (9).

0015

Table 10

Number and Percent Who Met One Criterion of Spanish Culture But Not a Second Criterion[a]

BEST COPY AVAILABLE

First Criterion	Survey Item or Criterion Number		Survey Item on Criterion Number—Second Criterion[a]											
			1	2	3	4	5	6	7	8	9	10	11	12	13
Spanish Spoken in Childhood Home	1	N	--	65	129	41	255	134	18	109	51	143	510	5	8
		%	--	7.7	11.8	4.8	32.8	16.9	1.5	13.3	5.4	15.8	58.8	0.5	0.8
Some Spanish Spoken Currently	2	N	28	--	129	41	230	130	28	116	59	124	478	12	19
		%	2.2	--	11.5	4.3	31.3	17.0	2.5	14.1	6.4	14.0	57.2	0.9	1.8
Considers Self as Spanish or Mexican-American	3	N	25	62	--	19	238	115	11	88	39	119	456	0	3
		%	2.7	7.9	--	2.8	34.2	16.1	1.1	12.7	4.8	14.7	58.8	0	0.4
Spanish Origin or Descent	4	N	59	95	140	--	268	138	34	119	48	143	528	14	24
		%	5.8	10.6	12.5	--	33.8	16.5	3.0	13.9	5.2	17.6	59.7	1.2	2.3
Either Parent Born in Spanish Speaking Country or Territory	5	N	32	43	119	29	--	23	10	79	42	101	351	1	5
		%	4.3	7.5	15.0	4.7	--	4.6	0.9	12.2	6.0	15.5	54.4	1	0.5
Two or More Grandparents Born in Spanish Speaking Country or Territory	6	N	53	86	137	41	165	--	16	85	53	150	465	4	11
		%	6.1	11.4	13.9	5.1	24.2	--	1.6	12.2	6.7	18.6	59.7	0.8	1.3
Either Parent Could Speak Spanish	7	N	64	110	161	62	274	142	--	113	67	180	555	14	16
		%	5.9	12.0	14.3	6.4	33.2	16.9	--	13.2	6.6	18.6	60.7	1.2	1.3
Two or More Grandparents Could Speak Spanish	8	N	51	94	133	44	242	109	11	--	49	152	485	4	12
		%	5.5	11.7	14.0	5.5	33.1	15.7	1.2	--	5.5	17.9	59.9	0.3	1.3
Considered Name to be Spanish or Mexican	9	N	76	119	161	54	277	152	46	120	--	169	528	29	39
		%	8.2	14.3	15.9	6.8	35.6	19.3	4.9	14.8	--	18.3	62.0	3.4	4.4
Had Most in Common With a Spanish Group	10	N	49	65	129	50	231	137	43	121	62	--	456	27	36
		%	5.6	8.9	13.4	6.5	33.5	19.3	4.9	15.9	7.1	--	59.8	3.0	4.4
Had Narrow Spanish First Name	11	N	10	15	57	15	75	48	14	50	25	49	--	5	5
		%	1.7	4.9	11.7	4.6	24.9	15.5	3.1	13.1	7.7	15.3	--	1.3	1.6
Spanish by Three or More Survey Criteria	12	N	62	106	161	53	281	141	25	117	61	178	562	--	19
		%	5.9	11.5	14.1	5.5	33.6	16.9	2.2	13.2	6.1	18.0	60.4	--	1.8
Spanish According to Discriminant Function[b]	13	N	48	96	147	46	268	132	11	110	55	170	545	2	--
		%	4.6	10.8	13.1	5.0	32.8	16.0	0.7	12.7	5.5	17.8	59.8	0.2	--

[a]The numbers are the unweighted number who met the first criterion but failed to meet the second, when both criteria were known. The denominator for each percentage is the number meeting the first criterion when both criteria are known. Prior to calculating the percentages, the cases are weighted to reflect the population shown in Table 1.

[b]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

0016

Table 11

Number and Percent Who Met One Criterion of Spanish Culture But Not a Second, From November 1969 Current Population Survey[a]

BEST COPY AVAILABLE

	Second Criterion					
	Spanish Origin		Spanish Mother Tongue		Spanish Current Language	
First Criterion	(Thousands)	Percent	(Thousands)	Percent	(Thousands)	Percent
Spanish Origin	--	--	2,872	31.1	4,732	51.3
Spanish Mother Tongue	342	5.1	--	--	2,131	31.8
Spanish Current Language	124	2.7	53	1.1	--	--

[a]The numbers are estimates based on the population surveyed.

These results, presented in Table 11, show the origin or descent criterion to be much more inclusive (and less specific) than the Spanish mother tongue criterion. Differences between the two criteria are much more exaggerated in the CPS study than present data would indicate. In the CPS, 31% of persons with Spanish origin or descent did not have a Spanish mother tongue, as compared with 6% in the present study. Such a difference is not impossible in view of the different populations involved. The present data exclude women and children and persons of non-Spanish surnames.

When the data are broken out by geographic area, as in Table 12, there is a variation in both the inclusiveness and specificity with geographic area. The Spanish mother tongue question (survey Item 1), for example, appears to be slightly more inclusive in the five southwestern states than the origin or descent question (survey Item 4), the two being about equal in specificity. In the New York, New Jersey, and Florida grouping, the two questions appear to be about equal in inclusiveness, with the Spanish mother tongue item having a slight edge in specificity. In the parts of the United States outside these areas, the origin or descent item is more inclusive, while the Spanish mother tongue item is more specific.

The "either parent spoke Spanish" criterion (Item 7) is more inclusive than the Census survey items (Items 1 and 4) in all geographic areas except Puerto Rico. This advantage is offset by a specificity which is poorer than that of the Census items in most areas and which might degrade seriously were the survey population not restricted to persons of Spanish surname.

All of the foregoing may be summarized as showing that alternatives to Spanish surname recognition are not perfect either. Persons who are Spanish by one criterion are not always Spanish by another. Moreover, the inclusiveness and specificity of a particular criterion varies with geographic area, and presumably by other population characteristics as well.

Despite these negative findings, alternatives to Spanish surname recognition nevertheless have the advantage of addressing whatever criterion is of most interest for a particular purpose. The false inclusions of Spanish surname recognition caused by culturally ambiguous names and acculturation are avoided.

Table 12

Number and Percent of Cases Who Met One Criterion of Spanish Culture But Not a Second, by Geographic Area

BEST COPY AVAILABLE

Geographic Area	First Survey Item or Criterion	Second Criterion[a]									
		Survey Item								Discriminant Function[b]	
		1		3		4		7			
		N	%	N	%	N	%	N	%	N	%
Five Southwestern States	1	--	--	26	5.7	24	5.0	4	0.8	3	0.6
	3	11	2.0	--	--	15	3.5	3	0.7	2	0.5
	4	17	3.2	22	5.0	--	--	6	1.0	5	0.7
	7	23	4.1	37	7.7	32	6.1	--	--	5	0.7
	Discriminant Function[b]	18	3.4	32	7.0	32	5.4	2	0.3	--	--
New York, New Jersey, or Florida	1	--	--	23	16.1	8	5.6	2	1.4	0	0
	3	4	3.2	--	--	2	1.7	3	2.4	1	0.8
	4	13	8.7	26	17.5	--	--	6	4.0	4	2.7
	7	15	9.5	35	22.3	14	8.9	--	--	3	1.9
	Discriminant Function[b]	12	7.6	32	20.5	11	7.1	2	1.3	--	--
Puerto Rico	1	--	--	54	30.4	0	0	7	3.9	0	0
	3	1	1.0	--	--	0	0	5	3.9	0	0
	4	3	1.7	56	30.9	--	--	7	3.8	0	0
	7	3	1.7	54	31.0	0	0	--	--	0	0
	Discriminant Function[b]	3	1.7	56	30.9	0	0	7	3.8	--	--
Elsewhere	1	--	--	26	23.2	9	6.8	5	3.5	5	3.5
	3	9	7.7	--	--	2	1.6	0	0	0	0
	4	26	17.7	36	27.3	--	--	15	10.4	15	10.4
	7	23	14.2	35	25.7	16	9.8	--	--	8	4.6
	Discriminant Function[b]	15	10.0	27	22.1	8	5.5	0	0	--	--

[a]The numbers are the unweighted number who met the first criterion but failed to meet the second, when both criteria were known. The denominator for each percentage is the number meeting the first criterion when both criteria are known. Prior to calculating the percentages, the cases are weighted to reflect the population shown in Table 1.

[b]The discriminant function was set to distinguish cases which were Spanish by three or more survey criteria.

SUMMARY AND CONCLUSIONS

Survey responses from 1,613 former servicemen with Spanish surnames were examined to evaluate Spanish surname recognition as a means of identifying persons of Latin-American birth or ancestry. Among the findings were the following:

(1) Outside certain areas Spanish surname recognition includes an unacceptably high proportion of persons who do not meet other criteria indicating Latin-American birth or ancestry.

(2) Within five southwestern states Spanish surname recognition includes increasing proportions of persons who could not be classified as Spanish at increasing educational and aptitude levels.

(3) These problems are diminished, but not eliminated, by using a Spanish surname list relatively free of culturally ambiguous names.

(4) Persons identified as Spanish by various criteria or combinations of criteria are often missed by a particular criterion. For example, persons who were Spanish by three or more survey criteria did not classify themselves as having Spanish origin or descent in 5% of the cases.

(5) The proportion of persons who meet one survey criterion but fail to meet a second varies with geographic area.

Despite the fact that alternatives to Spanish surname recognition share some of its idiosyncrasies, they are believed to be better. By addressing the criterion of interest directly, cases who no longer meet the criterion need not be included. This is not the case when a proxy variable is used, such as the presence of a Spanish surname.

0019

REFERENCES

1. Cabinet Committee on Opportunity for the Spanish Speaking. *Spanish Surnamed American College Graduates, 1971-1972*, 1972.
2. Ramirez, Henry M. "Americ"'s Spanish Speaking: A Profile," *Manpower*, September 1972.
3. Buechley, Robert W., "A Computer Program that Distinguishes Spanish Surnames," (unpublished).
4. Bureau of the Census. *1970 Census General Coding Procedures Manual*, Attachment J2.
5. Giesecke, G.L. *A Comparison of Computerized Techniques for Recognizing Spanish Names*, HumRRO Professional Paper 9-73, October 1973.
6. Morton, William E. *U.S. Bureau of the Census List of Spanish Surnames, Revised as of October 20, 1967*, University of Oregon Medical School, January 1968.
7. U.S. Air Force. *Airman Sample Survey*, USAF SCN 71-22 and 72-1, Data Services Center, HQ USAF, Washington, D.C.
8. Bureau of the Census. *Persons of Spanish Origin in the United States: March 1972 and 1971*, Series P-20, No. 250, U.S. Government Printing Office, Washington, D.C., April 1973.
9. Bureau of the Census. *Persons of Spanish Origin in the United States: November 1969*, Series P-20, No. 213, U.S. Government Printing Office, Washington, D.C., February 1971.

p 16 blank

0020

Appendix A

QUESTIONNAIRE

BEST COPY AVAILABLE

Mark one box for each of the following questions.

1. What language, other than English, was spoken in your home when you were a child?

(a) ☐ Spanish
(b) ☐ French
(c) ☐ German
(d) ☐ Other
Specify _____
(e) ☐ None, English only

2. How often do you speak Spanish among your friends and family?

(a) ☐ We speak only Spanish
(b) ☐ We speak mostly Spanish but also English
(c) ☐ We speak about equal amounts of Spanish and English
(d) ☐ We speak mostly English but also Spanish
(e) ☐ We speak English only
(f) ☐ Other, Specify _____

3. Which of the following would you consider yourself?

(a) ☐ Negro/Black
(b) ☐ Spanish or Mexican-American
(c) ☐ American Indian
(d) ☐ Oriental
(e) ☐ White
(f) ☐ Other
Specify _____

4. What is your origin or descent?

(a) ☐ Mexican
(b) ☐ Puerto Rican
(c) ☐ Cuban
(d) ☐ Central or South American
(e) ☐ Other Spanish
(f) ☐ None of these

5. Which of your parents was born in a Spanish-speaking country or territory?

(a) ☐ Neither
(b) ☐ Mother
(c) ☐ Father
(d) ☐ Both

6. How many of your grandparents were born in a Spanish-speaking country or territory?

(a) ☐ None
(b) ☐ One
(c) ☐ Two
(d) ☐ Three
(e) ☐ Four

PLEASE TURN OVER

0021

BEST COPY AVAILABLE

7. Which of your parents can (or could) carry on a long conversation in Spanish?

(a) ☐ Neither
(b) ☐ Mother
(c) ☐ Father
(d) ☐ Both

8. How many of your grandparents can (or could) carry on a long conversation in Spanish?

(a) ☐ None
(b) ☐ One
(c) ☐ Two
(d) ☐ Three
(e) ☐ Four

9. What is the origin or nationality of your last name?

(a) ☐ Spanish
(b) ☐ Portuguese
(c) ☐ Italian
(d) ☐ Yiddish
(e) ☐ Filipino
(f) ☐ American Indian
(g) ☐ Mexican
(h) ☐ Polynesian
(i) ☐ Arabic
(j) ☐ African
(k) ☐ Greek
(l) ☐ Polish
(m) ☐ Czech
(n) ☐ Hungarian
(o) ☐ Russian
(p) ☐ German
(q) ☐ French
(r) ☐ Dutch
(s) ☐ Scandinavian
(t) ☐ English
(u) ☐ Scottish
(v) ☐ Irish
(w) ☐ Chinese
(x) ☐ Japanese
(y) ☐ Not known
(z) ☐ Other
Specify _____

10. Besides being an American, with which of the following people who live in the United States do you have the most in common?

(a) ☐ Mexican Americans
(b) ☐ Puerto Ricans
(c) ☐ Cubans
(d) ☐ Dominicans
(e) ☐ Latin Americans
(f) ☐ Central Americans
(g) ☐ South Americans
(h) ☐ Whites or Anglos
(i) ☐ Blacks or Negroes
(j) ☐ Spaniards
(k) ☐ Spanish Americans
(l) ☐ Mexicans
(m) ☐ Indians
(n) ☐ Chicanos
(o) ☐ Hispanos
(p) ☐ Other
Specify _____

Thank you for completing the questionnaire. Please return it to us in the accompanying envelope. If there are any comments you would like to make to help clarify your answers, please do so here.

Unclassified

BEST COPY AVAILABLE

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: HumRRO-PP-22-74	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): THE USE OF SPANISH SURNAMES AS A MEANS OF IDENTIFYING LATINOS IN THE UNITED STATES AND PUERTO RICO	5. TYPE OF REPORT & PERIOD COVERED: Professional Paper 6. PERFORMING ORG. REPORT NUMBER: Professional Paper 22-74
7. AUTHOR(s): G. Lee Giesecke	8. CONTRACT OR GRANT NUMBER(s): MDA 903-74-C-0005
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Human Resources Research Organization (HumRRO) 300 North Washington Street Alexandria, Virginia 22314	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Directorate for Manpower Research Office of the Assistant Secretary of Defense (Manpower and Reserve Affairs)	12. REPORT DATE: December 1974 13. NUMBER OF PAGES: 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES: Research performed by HumRRO Division No. 7 (Social Science)	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Classification procedures Coding techniques Computer technology Ethnic classification Name classification Spanish surnames Questionnaire survey	
20. ABSTRACT (Continue on reverse side if necessary and identify by block number): In this paper, the use of Spanish surnames as a way to identify Latinos (persons living in the United States with Latin-American birth or ancestry) is examined. The problems attendant upon this method are discussed, along with the question of whether there are more attractive alternatives. Data are from a questionnaire mailed to a sample of 5,019 former servicemen, meeting various criteria, including a Spanish surname.	

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)